



University of Pennsylvania  
**ScholarlyCommons**

---

Departmental Papers (ASC)

Annenberg School for Communication

---

1-25-2011

# Computing Krippendorff's Alpha-Reliability

Klaus Krippendorff

*University of Pennsylvania*, [kkrippendorff@asc.upenn.edu](mailto:kkrippendorff@asc.upenn.edu)

Follow this and additional works at: [http://repository.upenn.edu/asc\\_papers](http://repository.upenn.edu/asc_papers)



Part of the [Communication Commons](#)

---

## Recommended Citation

Krippendorff, K. (2011). Computing Krippendorff's Alpha-Reliability. Retrieved from [http://repository.upenn.edu/asc\\_papers/43](http://repository.upenn.edu/asc_papers/43)

Postprint version.

This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/asc\\_papers/43](http://repository.upenn.edu/asc_papers/43)

For more information, please contact [libraryrepository@pobox.upenn.edu](mailto:libraryrepository@pobox.upenn.edu).

---

# Computing Krippendorff's Alpha-Reliability

## **Abstract**

Krippendorff's alpha ( $\alpha$ ) is a reliability coefficient developed to measure the agreement among observers, coders, judges, raters, or measuring instruments drawing distinctions among typically unstructured phenomena or assign computable values to them.  $\alpha$  emerged in content analysis but is widely applicable wherever two or more methods of generating data are applied to the same set of objects, units of analysis, or items and the question is how much the resulting data can be trusted to represent something real.

## **Disciplines**

Communication | Social and Behavioral Sciences

## **Comments**

Postprint version.

# Computing Krippendorff's Alpha-Reliability

Klaus Krippendorff  
kkrippendorff@asc.upenn.edu  
2011.1.25

**Krippendorff's alpha ( $\alpha$ )** is a reliability coefficient developed to measure the agreement among observers, coders, judges, raters, or measuring instruments drawing distinctions among typically unstructured phenomena or assign computable values to them.  $\alpha$  emerged in content analysis but is widely applicable wherever two or more methods of generating data are applied to the same set of objects, units of analysis, or items and the question is how much the resulting data can be trusted to represent something real.

$\alpha$ 's general form is:

$$\alpha = 1 - \frac{D_o}{D_e}$$

where  $D_o$  is the observed disagreement among values assigned to units of analysis:

$$D_o = \frac{1}{n} \sum_c \sum_k o_{ck \text{ metric}} \delta_{ck}^2$$

and  $D_e$  is the disagreement one would expect when the coding of units is attributable to chance rather than to the properties of these units:

$$D_e = \frac{1}{n(n-1)} \sum_c \sum_k n_c \cdot n_k \text{ metric} \delta_{ck}^2$$

The arguments in the two disagreement measures,  $o_{ck}$ ,  $n_c$ ,  $n_k$  and  $n$ , refer to the frequencies of values in coincidence matrices, to be defined below.

Algebraically, when observers agree perfectly, observed disagreement  $D_o=0$  and  $\alpha=1$ , which indicates perfect reliability. When observers agree as if chance had produced the results,  $D_o=D_e$  and  $\alpha=0$ , which indicates the absence of reliability.  $\alpha=0$  occurs when observers are unable to distinguish among units or assign values to them drawn randomly from a collective estimate of the population of data. To rely on data generated by any method,  $\alpha$  needs to be far from these two extreme conditions, ideally  $\alpha=1$ . For reliability considerations,  $\alpha$ 's range is:

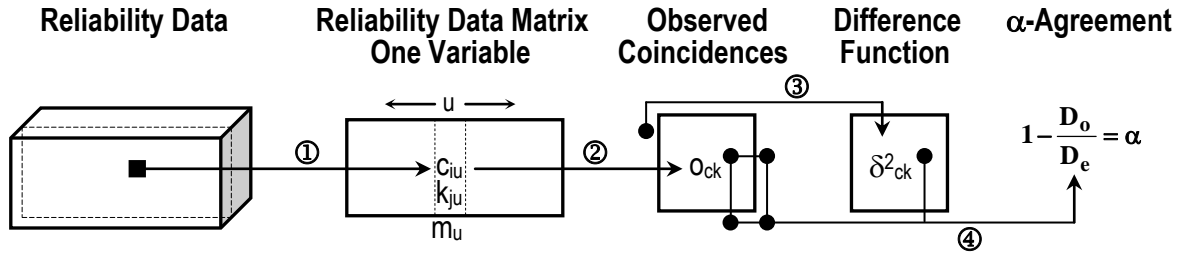
$$1 \geq \alpha \geq 0 \begin{cases} - \text{Systematic disagreement} \\ \pm \text{Sampling errors} \end{cases}$$

Unlike other specialized coefficients,  $\alpha$  is a generalization of several known reliability indices. It enables researchers to judge a variety of data with the same reliability standard.  $\alpha$  applies to:

- Any number of observers, not just two
- Any number of categories, scale values, or measures
- Any metric or level of measurement (nominal, ordinal, interval, ratio, and more)
- Incomplete or missing data
- Large and small sample sizes alike, not requiring a minimum

$\alpha$  evaluates reliability one variable at a time. It offers other analytical possibilities not presented here.

**Reliability data** duplicate the process of generating data whose reliability is in question. Given such data,  $\alpha$ -reliability can be computed in four computational steps, graphed below.



These four computational steps will be defined and illustrated with four kinds of data of increasing complexity:

- |  |               |
|--|---------------|
| <b>A. Binary or dichotomous data, two observers, no missing data</b> | <b>Page 2</b> |
| <b>B. Nominal data, two observers, no missing data</b>               | <b>Page 3</b> |
| <b>C. Nominal data, any number of observers, missing data</b>        | <b>Page 4</b> |
| <b>D. All metrics, any number of observers, missing data</b>         | <b>Page 5</b> |

Finally,

- E. A general computational form** is presented, bypassing coincidence matrices **Page 9**

## A. Binary or dichotomous data, two observers, no missing data

① Construct a **reliability data matrix**; here, a **2 observers-by- $N$  units** matrix, containing  $2N$  values, **c** and **k**:

Units:	<b>1</b>	<b>2</b>	...	<b><math>u</math></b>	...	<b><math>N</math></b>	
Observers:	<b><math>i</math>:</b>	$c_{i1}$	$c_{i2}$	...	$c_{iu}$	...	$c_{iN}$
	<b><math>j</math>:</b>	$c_{j1}$	$c_{j2}$	...	$c_{ju}$	...	$c_{jN}$

For example, when two observers judge  $N=10$  units, the 2-by-10 data matrix contains 20 values:

Items judged:	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>
<b>Meg:</b>	0	1	0	0	0	0	0	0	1	0
<b>Owen:</b>	1	1	1	0	0	1	0	0	0	0

② Tabulate **coincidences within units**. Coincidence matrices account for *all values* contained in a reliability data matrix. They differ from the familiar contingency matrices, which account for *units* in two dimensions, not values. The importance of this difference becomes apparent in C.

Into a 2-by-2 coincidence matrix, units are entered twice, once as **c-k** pairs and once as **k-c** pairs. In the example, unit **1** is entered as a **0-1** pair of values *and* as a **1-0** pair of values. Unit **2** is entered as two **1-1** pairs of values, etc.:

Values:	<b>0</b>	<b>1</b>	
<b>0</b>	$o_{00}$	$o_{01}$	$n_0$
<b>1</b>	$o_{10}$	$o_{11}$	$n_1$
Number of Values:	$n_0$	$n_1$	$n=2N$

	<b>0</b>	<b>1</b>	
<b>0</b>	10	4	14
<b>1</b>	4	2	6
	14	6	20

Accordingly,  $o_{00}$  represents the ten **0-0** pairs within units **4, 5, 7, 8, and 10**.  $o_{01}$  represents the four **0-1** pairs in units **1, 3, 6, and 9**, and  $o_{10}$  represents the four **1-0** pairs in the four same units.  $o_{11}$  represents the two **1-1** pairs found only in unit **2**.  $n_0=14$  is the number of **0s** in the reliability data matrix,  $n_1=6$  is the number of **1s**, and  $n=2N=20$  is the total number of values paired.

For these binary data, mismatching coincidences occur in two cells  $o_{01}$ ,  $o_{10}$  of equal frequency, 4.

③ skip

④ Compute  **$\alpha$ -reliability** (most simple form): 
$$\text{binary } \alpha = 1 - \frac{D_o}{D_e} = 1 - (n-1) \frac{o_{01}}{n_0 \cdot n_1}$$

In the example:

$$\text{binary } \alpha = 1 - (20-1) \frac{4}{14 \cdot 6} = 0.095$$

## B. Nominal data, two observers, no missing data

① Construct a **reliability data matrix** – just as in A above. For a 2-by-12 example:

Items judged: **1 2 3 4 5 6 7 8 9 10 11 12**

<b>Ben:</b>	<i>a</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>d</i>	<i>c</i>	<i>c</i>	<i>c</i>	<i>e</i>	<i>d</i>	<i>d</i>	<i>a</i>
<b>Gerry:</b>	<i>b</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>b</i>	<i>c</i>	<i>c</i>	<i>c</i>	<i>e</i>	<i>d</i>	<i>d</i>	<i>d</i>

② Tabulate **coincidences within units**. The general form of a coincidence matrix and one with frequencies from the above example entered are:

Values:	<b>1 . k . .</b>	
<b>1</b>	$o_{11} . o_{1k} . .$	$n_1$
.	. . . .	.
.	. . . .	.
<b>c</b>	$o_{c1} . o_{ck} . .$	$n_c = \sum_k o_{ck}$
.	. . . .	.
	$n_1 . n_k . .$	$n = \sum_c \sum_k o_{ck}$

	<b>a b c d e</b>	
<b>a</b>	2 1 . 1 .	4
<b>b</b>	1 4 . 1 .	6
<b>c</b>	. . 6 . .	6
<b>d</b>	1 1 . 4 .	6
<b>e</b>	. . . . 2	2
	4 6 6 6 2	24

Where  $o_{ck} = \sum_u$  Number of **c-k** pairs in unit  $u$  specifically:

$o_{ab} = 1$  **a-b** pair in unit 1  
 $o_{ba} = 1$  **b-a** pair in unit 1  
 $o_{aa} = 2$  **a-a** pairs in unit 2  
 $o_{bb} = 4 = 2$  **b-b** pairs in unit 3  
 + 2 **b-b** pairs in unit 4  
 and so forth.

$n_a=4$  is the number of **as**

$n_b=6$  is the number of **bs**

and so forth.

$n=24$  is the total number of values  
 for two observers:  $n = 2N$

③ skip

④ Compute  **$\alpha$ -reliability** (most simple form):

$$\text{nominal } \alpha = 1 - \frac{D_o}{D_e} = \frac{A_o - A_e}{1 - A_e} = \frac{(n-1) \sum_c o_{cc} - \sum_c n_c (n_c - 1)}{n(n-1) - \sum_c n_c (n_c - 1)}$$

Wherein  $A_o$  is the percent of observed matches in units  $u$  and  $A_e$  is the percent of matches obtainable by chance. The computational form further simplifies the needed computations.

In the example:

$$\alpha_{\text{nominal}} = \frac{(24-1)(2+4+6+4+2) - [4(4-1) + 6(6-1) + 6(6-1) + 6(6-1) + 2(2-1)]}{24(24-1) - [4(4-1) + 6(6-1) + 6(6-1) + 6(6-1) + 2(2-1)]} = 0.692$$

## C. Nominal data, any number of observers, missing data

① Construct a **reliability data matrix** – just as in A and in B above, but for  $m$  observers:

Units $u$ :	<b>1</b>	<b>2</b>	.	.	.	$u$	.	.	.	.	.	$N$
Observers: <b>1</b>	$c_{11}$	$c_{12}$	.	.	.	$c_{1u}$	.	.	.	.	.	$c_{1N}$
<b><math>i</math></b>	$c_{i1}$	$c_{i2}$	.	.	.	$c_{iu}$	.	.	.	.	.	$c_{iN}$
<b><math>j</math></b>	$c_{j1}$	$c_{j2}$	.	.	.	$c_{ju}$	.	.	.	.	.	$c_{jN}$
.	.	.	.	.	.	.	.	.	.	.	.	.
<b><math>m</math></b>	$c_{m1}$	$c_{m2}$	.	.	.	$c_{mu}$	.	.	.	.	.	$c_{mN}$

Number of observers valuing  $u$ :  $m_1$   $m_2$  . . .  $m_u$  . . .  $m_N$

When data are missing, this matrix contains less than  $mN$  entries and  $m_u$  is variable.

For example, a 4 observers-by-12 units reliability data matrix:

Units $u$ :	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>
Observer <b>A</b> :	1	2	3	3	2	1	4	1	2	.	.	.
Observer <b>B</b> :	1	2	3	3	2	2	4	1	2	5	.	3
Observer <b>C</b> :	.	3	3	3	2	3	4	2	2	5	1	.
Observer <b>D</b> :	1	2	3	3	2	4	4	1	2	5	1	.
Number $m_u$ of values in unit $u$ :	3	4	4	4	4	4	4	4	4	3	2	1

Note that 7 out of the 48 possible values in this matrix are missing.  $m_u$  varies from 1 to 4.

② Tabulate **coincidences within units**. The coincidence matrix appears as in B:

Values:	<b>1</b>	.	<b>k</b>	.	.
<b>1</b>	$o_{11}$	$o_{1k}$	.	.	$n_1$
.	.	.	.	.	.
.	.	.	.	.	.
<b>c</b>	$o_{c1}$	$o_{ck}$	.	.	$n_c = \sum_k o_{ck}$
.	.	.	.	.	.
	$n_1$	$n_k$	.	.	$n = \sum_c \sum_k n_{ck}$

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	
<b>1</b>	7	4/3	1/3	1/3	.	9
<b>2</b>	4/3	10	4/3	1/3	.	13
<b>3</b>	1/3	4/3	8	1/3	.	10
<b>4</b>	1/3	1/3	1/3	4	.	5
<b>5</b>	.	.	.	.	3	3
	9	13	10	5	3	40

Unlike in the two-observer case in B:  $o_{ck} = \sum_u \frac{\text{Number of c-k pairs in unit } u}{m_u - 1}$

Note that each unit contains  $m_u(m_u-1)$  coincidences. A coincidence matrix accounts for all pairs of values found in  $u$ . Unit **1** contains  $3(3-1)=6$  pairs of matching **1s**. It contributes

6/(3-1)=3 to the  $o_{11}$  cell, one for each value. Unit **2** contains 4(4-1)=12 pairs, 6 matching **2-2** pairs, 3 mismatching **2-3** pairs, and 3 mismatching **3-2** pairs. It adds 6/(4-1)=2 to  $o_{22}$ , 3/(4-1)=1 to  $o_{23}$ , 1 to  $o_{32}$ , and 4 to the total  $n$ , thus fully accounting for its 4 values. Unit **6** contains 4·(4-1)=12 pairs of mismatching values, each adds 1/(4-1)=1/3 to a different cell. The lone value **3** in unit **12** affords no comparisons, 1·(1-1)=0 pairs and does not add to this account. Thus, the margins of coincidence matrices do not represent all values that occur in a reliability data matrix, only those that can be paired within units, here  $n=40$  pairable values over all units.

③ Skip

④ Compute  **$\alpha$ -reliability** – just as in B  $\alpha_{\text{nominal}} = 1 - \frac{D_o}{D_e} = \frac{A_o - A_e}{1 - A_e} = \frac{(n-1)\sum_c o_{cc} - \sum_c n_c(n_c - 1)}{n(n-1) - \sum_c n_c(n_c - 1)}$

In the example:

$$\alpha_{\text{nominal}} = \frac{(40-1)(7+10+8+4+3) - [9(9-1) + 13(13-1) + 10(10-1) + 5(5-1) + 3(3-1)]}{40(40-1) - [9(9-1) + 13(13-1) + 10(10-1) + 5(5-1) + 3(3-1)]} = 0.743$$

## D. Any metric, any number of observers, missing data

① Construct a **reliability data matrix** – just as in C

② Tabulate **coincidences** within units – just as in C

③ Insert the **difference function**  $\delta_{ck}^2$  that is appropriate to the metric of the given data into the two disagreements  $D_o$  and  $D_e$  defined in the beginning of this document.

Note that  $\alpha$  accounts for different metrics or levels of measurement by weighing the observed and expected coincidences by the squared difference between the coinciding values. Differences can be expressed as mathematical functions and in the form of a table. The latter makes their relative magnitudes transparent. Interval and ratio metric differences are functions of the values being paired. Ordinal differences depend on their frequencies of using values. And nominal differences are added here to generalize step ④.

- **Nominal metric differences** – Two values either match, or they do not:

Nominal categories, names:		<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>
$\alpha_{\text{nominal}} \delta_{ck}^2 = \begin{cases} 0 & \text{iff } c = k \\ 1 & \text{iff } c \neq k \end{cases}$	<i>a</i>	0	1	1	1	1	1
	<i>b</i>	1	0	1	1	1	1
	<i>c</i>	1	1	0	1	1	1
	<i>d</i>	1	1	1	0	1	1
	<i>e</i>	1	1	1	1	0	1
	<i>f</i>	1	1	1	1	1	0

- **Ordinal metric differences** – Values have the meaning of ranks and differences between ranks depend on how many ranks they are apart from each other. For example, with frequencies from data in C (and one unused rank added to show that it does not matter):

$$\text{ordinal } \delta_{ck}^2 = \left( \sum_{g=c}^{g=k} n_g - \frac{n_c + n_k}{2} \right)^2$$

<b>Ranks:</b>	<b>1<sup>st</sup></b>	<b>2<sup>nd</sup></b>	<b>3<sup>rd</sup></b>	<b>4<sup>th</sup></b>	<b>5<sup>th</sup></b>	<b>6<sup>th</sup></b>
<b>1<sup>st</sup></b>	0	11 <sup>2</sup>	22.5 <sup>2</sup>	30 <sup>2</sup>	32.5 <sup>2</sup>	34 <sup>2</sup>
<b>2<sup>nd</sup></b>	121	0	11.5 <sup>2</sup>	19 <sup>2</sup>	21.5 <sup>2</sup>	23 <sup>2</sup>
<b>3<sup>rd</sup></b>	506	132	0	7.5 <sup>2</sup>	10 <sup>2</sup>	11.5 <sup>2</sup>
<b>4<sup>th</sup></b>	900	361	56	0	2.5 <sup>2</sup>	4 <sup>2</sup>
<b>5<sup>th</sup></b>	992	462	100	6.3	0	1.5 <sup>2</sup>
<b>6<sup>th</sup></b>	1156	529	132	16	2.3	0

$n_{1st} = 9$		
$n_{2nd} = 13$		
$n_{3rd} = 10$		
$n_{4th} = 5$		
$n_{5th} = 0$		
$n_{6th} = 3$		

↑

↑

$\delta_{1^{st}, 3^{rd}} = 22.5$   
 $\delta_{4^{th}, 6^{th}} = 4$

Ordinal metric differences may be standardized:  $0 \leq \text{ordinal } \delta_{ck}^2 \leq 1$  by:

$$\text{ordinal } \delta_{ck}^2 = \left( \frac{\sum_{g=c}^{g=k} n_g - \frac{n_c + n_k}{2}}{n - \frac{n_{c_{max}} + n_{c_{min}}}{2}} \right)^2$$

where  $c_{max}$  is the largest and  $c_{min}$  the smallest rank among all ranks used. Standardization does not affect  $\alpha$ , however.

- **Interval metric differences** – Values differ algebraically:

$$\text{interval } \delta_{ck}^2 = (c - k)^2$$

Interval values:	<b>-1</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
<b>-1</b>	0	1 <sup>2</sup>	2 <sup>2</sup>	3 <sup>2</sup>	4 <sup>2</sup>	5 <sup>2</sup>
<b>0</b>	1	0	1 <sup>2</sup>	2 <sup>2</sup>	3 <sup>2</sup>	4 <sup>2</sup>
<b>1</b>	4	1	0	1 <sup>2</sup>	2 <sup>2</sup>	3 <sup>2</sup>
<b>2</b>	9	4	1	0	1 <sup>2</sup>	2 <sup>2</sup>
<b>3</b>	16	9	4	1	0	1 <sup>2</sup>
<b>4</b>	25	16	9	4	1	0

Interval metric differences may be standardized as well:  $0 \leq \text{interval } \delta_{ck}^2 \leq 1$  by:

$$\text{interval } \delta_{ck}^2 = \left( \frac{c - k}{c_{max} - c_{min}} \right)^2$$

where  $c_{max}$  is the largest and  $c_{min}$  is the smallest value occurring in the data. This standardization does not affect  $\alpha$  either.



- **Ratio metric differences** – Algebraic differences between two values are expressed relative to an absolute zero point. They are proportional to the magnitude of their values:

Ratio values:	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
$\delta_{ck}^2 = \left( \frac{c-k}{c+k} \right)^2$	0	$(\frac{1}{2})^2$	$(\frac{2}{3})^2$	$(\frac{3}{4})^2$	$(\frac{4}{5})^2$	$(\frac{5}{6})^2$
	<b>1</b>	1	0	$(\frac{1}{3})^2$	$(\frac{2}{4})^2$	$(\frac{3}{5})^2$
	<b>2</b>	1	.11	0	$(\frac{1}{5})^2$	$(\frac{2}{6})^2$
	<b>3</b>	1	.25	.04	0	$(\frac{1}{7})^2$
	<b>4</b>	1	.36	.11	.02	0
	<b>5</b>	1	.44	.18	.06	.01

- **Circular metric differences** – Shortest differences between any two values on a circular scale with arbitrary endpoints but a fixed circumference U = the number of equal intervals on a circle.

Circular values:	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
With the sine function expressed in degrees:	0	.25	.75	1	.75	.25
$\delta_{ck}^2 = \left( \sin \left[ 180 \frac{c-k}{U} \right] \right)^2$	<b>1</b>	.25	0	.25	.75	1
	<b>2</b>	.75	.25	0	.25	.75
With the sine function expressed in radian:	<b>3</b>	1	.75	.25	0	.25
$\delta_{ck}^2 = \left( \sin \left[ \pi \frac{c-k}{U} \right] \right)^2$	<b>4</b>	.75	1	.75	.25	0
	<b>5</b>	.25	.75	1	.75	.25

- **Bipolar metric differences** – Algebraic differences are expressed relative to the two endpoints,  $c_{\min}$  and  $c_{\max}$ , of the scale. Near the center, a bipolar metric behaves like an interval metric and near the poles it behaves like a ratio metric.

Bipolar values:	<b>-2</b>	<b>-1</b>	<b>0</b>	<b>1</b>	<b>2</b>
$\delta_{ck}^2 = \frac{(c-k)^2}{(c+k-2c_{\min})(2c_{\max}-c-k)}$	-2	0	$\frac{1^2}{1.7}$	$\frac{2^2}{2.6}$	$\frac{3^2}{3.5}$
	<b>-1</b>	.143	0	$\frac{1^2}{3.5}$	$\frac{2^2}{4.4}$
	<b>0</b>	.333	.067	0	$\frac{1^2}{5.3}$
	<b>1</b>	.600	.250	.067	0
	<b>2</b>	1.00	.600	.333	.143

④ Compute  **$\alpha$ -reliability** (the computationally most efficient form):

$$\alpha_{\text{metric}} = 1 - \frac{D_o}{D_e} = 1 - (n-1) \frac{\sum_c \sum_{k>c} o_{ck \text{ metric}} \delta_{ck}^2}{\sum_c \sum_{k>c} n_c \cdot n_k \text{ metric} \delta_{ck}^2}$$

Note that the sums in this form enumerate only one of the two symmetrical off-diagonal triangles of a coincidence matrix. Its entries,  $o_{ck}$  as well as the products  $n_c n_k$  are weighted by an appropriately chosen difference function  $\text{metric} \delta_{ck}^2$ .

Computations are illustrated with the numerical data in C above, interpreted as ordinal, interval and ratio data respectively. Zero frequencies appear as 0 in the list of multiplications:

With data in C as **ordinal data**:

$$\alpha_{\text{ordinal}} = 1 - (n-1) \frac{\sum_c \sum_{k>c} o_{ck \text{ ordinal}} \delta_{ck}^2}{\sum_c \sum_{k>c} n_c \cdot n_k \text{ ordinal} \delta_{ck}^2}$$

In the Example:

$$\alpha_{\text{ordinal}} = 1 - (40-1) \frac{\frac{4}{3}11^2 + \frac{1}{3}22.5^2 + \frac{1}{3}30^2 + 0 + \frac{4}{3}11.5^2 + \frac{1}{3}19^2 + 0 + \frac{1}{3}7.5^2 + 0 + 0}{9 \cdot 13 \cdot 11^2 + 9 \cdot 10 \cdot 22.5^2 + 9 \cdot 5 \cdot 30^2 + 9 \cdot 3 \cdot 34^2 + 13 \cdot 10 \cdot 11.5^2 + 13 \cdot 5 \cdot 19^2 + 13 \cdot 3 \cdot 23^2 + 10 \cdot 5 \cdot 7.5^2 + 10 \cdot 3 \cdot 11.5^2 + 5 \cdot 3 \cdot 4^2} = 0.815$$

With data in C as **interval data**:

$$\alpha_{\text{interval}} = 1 - (n-1) \frac{\sum_c \sum_{k>c} o_{ck \text{ interval}} \delta_{ck}^2}{\sum_c \sum_{k>c} n_c \cdot n_k \text{ interval} \delta_{ck}^2}$$

In the example:

$$\alpha_{\text{interval}} = 1 - (40-1) \frac{\frac{4}{3}1^2 + \frac{1}{3}2^2 + \frac{1}{3}3^2 + 0 + \frac{4}{3}1^2 + \frac{1}{3}2^2 + 0 + \frac{1}{3}1^2 + 0 + 0}{9 \cdot 13 \cdot 1^2 + 9 \cdot 10 \cdot 2^2 + 9 \cdot 5 \cdot 3^2 + 9 \cdot 3 \cdot 4^2 + 13 \cdot 10 \cdot 1^2 + 13 \cdot 5 \cdot 2^2 + 13 \cdot 3 \cdot 3^2 + 10 \cdot 5 \cdot 1^2 + 10 \cdot 3 \cdot 2^2 + 5 \cdot 3 \cdot 1^2} = 0.849$$

With data in C as **ratio data**:

$$\alpha_{\text{ratio}} = 1 - (n-1) \frac{\sum_c \sum_{k>c} o_{ck \text{ ratio}} \delta_{ck}^2}{\sum_c \sum_{k>c} n_c \cdot n_k \text{ ratio} \delta_{ck}^2}$$

In the example:

$$\alpha_{\text{ratio}} = 1 - (40-1) \frac{\frac{4}{3} \frac{1^2}{3^2} + \frac{1}{3} \frac{2^2}{4^2} + \frac{1}{3} \frac{3^2}{5^2} + 0 + \frac{4}{3} \frac{1^2}{5^2} + \frac{1}{3} \frac{2^2}{6^2} + 0 + \frac{1}{3} \frac{1^2}{7^2} + 0 + 0}{9 \cdot 13 \cdot \frac{1^2}{3^2} + 9 \cdot 10 \cdot \frac{2^2}{4^2} + 9 \cdot 5 \cdot \frac{3^2}{5^2} + 9 \cdot 3 \cdot \frac{4^2}{6^2} + 13 \cdot 10 \cdot \frac{1^2}{5^2} + 13 \cdot 5 \cdot \frac{2^2}{6^2} + 13 \cdot 3 \cdot \frac{3^2}{7^2} + 10 \cdot 5 \cdot \frac{1^2}{7^2} + 10 \cdot 3 \cdot \frac{2^2}{8^2} + 5 \cdot 3 \cdot \frac{1^2}{9^2}} = 0.797$$

## E. A general computational form, bypassing coincidence matrices:

① Start from a **reliability data matrix** as in C above:

Units <b>u:</b>	<b>1</b>	<b>2</b>	.	.	.	<b>u</b>	.	.	.	.	.	<b>N</b>
Observers: <b>1</b>	$c_{11}$	$c_{12}$	.	.	.	$c_{1u}$	.	.	.	.	.	$c_{1N}$
<b>i</b>	$c_{i1}$	$c_{i2}$	.	.	.	$c_{iu}$	.	.	.	.	.	$c_{iN}$
<b>j</b>	$c_{j1}$	$c_{j2}$	.	.	.	$c_{ju}$	.	.	.	.	.	$c_{jN}$
.	.	.	.	.	.	.	.	.	.	.	.	.
<b>m</b>	$c_{m1}$	$c_{m2}$	.	.	.	$c_{mu}$	.	.	.	.	.	$c_{mN}$

When data are missing, this matrix will contain fewer than  $mN$  values.

For the 4 observers-by-12 units example of reliability data used in C and D above:

Units <b>u:</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>
Observers: <b>A:</b>	1	2	3	3	2	1	4	1	2	.	.	.
<b>B:</b>	1	2	3	3	2	2	4	1	2	5	.	3
<b>C:</b>	.	3	3	3	2	3	4	2	2	5	1	.
<b>D:</b>	1	2	3	3	2	4	4	1	2	5	1	.

Note that

- Out of the  $mN=4 \cdot 12=48$  possible values in this matrix, 7 are missing.
- The value 3, assigned by observer B to the 12<sup>th</sup> unit cannot be paired with non existing values in that unit, cannot contribute to observed agreements or disagreements, drops out when constructing a coincidence matrix, and has to be ignored.

Thus, this matrix contains a total of  $n_{..}=40$  pairable values.

Instead of ② and ③,

enumerate the values found in units and create a **values-by-units matrix**:

Units:	<b>1</b>	<b>2</b>	.	.	.	<b>u</b>	.	.	.	.	.	.
Values: <b>1</b>	$n_{11}$	$n_{21}$	.	.	.	$n_{u1}$	.	.	.	.	.	$n_{.1}$
.	.	.	.	.	.	.	.	.	.	.	.	.
<b>c</b>	$n_{1c}$	$n_{2c}$	.	.	.	$n_{uc}$	.	.	.	.	.	$n_{.c}$
<b>k</b>	$n_{1k}$	$n_{2k}$	.	.	.	$n_{uk}$	.	.	.	.	.	$n_{.k}$
.	.	.	.	.	.	.	.	.	.	.	.	.
Totals:	$n_{.1}$	$n_{.2}$	.	.	.	$n_{u.}$	.	.	.	.	.	$n_{..}$

Where  $n_{uc}$  = the number of values  $c$  assigned to unit  $u$ .  $n_{uc} \leq m$  observers.  $n_{uk}$  by analogy

$n_{u.} = \sum_c n_{uc}$  = the number of values assigned to unit  $u$

$n_{.c} = \sum_{u/n_{u.} \geq 2} n_{uc}$  = the number of pairable values  $c$  occurring in the reliability data (omitting all units with lone or no values:  $n_{u.} \leq 1$ )

$n_{..} = \sum_{u/n_{u.} \geq 2} n_{u.}$  = the total number of all pairable values in the reliability data (omitting all units with lone or no values:  $n_{u.} \leq 1$ );  $n_{..} \leq mN$

For the above example

Units $u$ :	1	2	3	4	5	6	7	8	9	10	11	12	
Values $c, k$ : 1	3	0	0	0	0	1	0	3	0	0	2	0	$9 = n_{.1}$
2	0	3	0	0	4	1	0	1	4	0	0	0	$13 = n_{.2}$
3	0	1	4	4	0	1	0	0	0	0	0	1	$10 = n_{.3}$
4	0	0	0	0	0	1	4	0	0	0	0	0	$5 = n_{.4}$
5	0	0	0	0	0	0	0	0	0	3	0	0	$3 = n_{.5}$
Totals $n_{u.}$ :	3	4	4	4	4	4	4	4	4	4	2	1	$40 = n_{..}$

Note that the marginal sum  $n_{.3}$  of pairable values 3 omits the only lone value in unit 12,  $n_{12.}=1$ .

④ **Compute  $\alpha$**  with the appropriate metric difference function as defined in D above:

$${}_{metric}\alpha = 1 - \frac{D_o}{D_e} = 1 - (n_{..} - 1) \frac{\sum_u \frac{1}{n_{u.} - 1} \sum_c \sum_{k>c} n_{uc} n_{uk} {}_{metric}\delta_{ck}^2}{\sum_c \sum_{k>c} n_{.c} n_{.k} {}_{metric}\delta_{ck}^2}$$

If the above example consists of **nominal data** (perfect agreement seen as 0 disagreement):

$${}_{nominal}\alpha = 1 - (40 - 1) \frac{0 + \frac{3 \cdot 1}{4 - 1} + 0 + 0 + 0 + 6 \frac{1 \cdot 1}{4 - 1} + 0 + \frac{3 \cdot 1}{4 - 1} + 0 + 0 + 0}{9 \cdot 13 + 9 \cdot 10 + 9 \cdot 5 + 9 \cdot 3 + 13 \cdot 10 + 13 \cdot 5 + 13 \cdot 3 + 10 \cdot 5 + 10 \cdot 3 + 5 \cdot 3} = 0.743$$

If the above example consists of **interval data** (perfect agreement seen as 0 disagreement):

$${}_{interval}\alpha = 1 - (40 - 1) \frac{0 + \frac{3 \cdot 1 \cdot 1^2}{4 - 1} + 0 + 0 + 0 + \frac{1 \cdot 1 \cdot 1^2 + 1 \cdot 1 \cdot 2^2 + 1 \cdot 1 \cdot 3^2 + \dots + 1 \cdot 1 \cdot 1^2}{4 - 1} + 0 + \frac{3 \cdot 1 \cdot 1^2}{4 - 1} + 0 + 0 + 0}{9 \cdot 13 \cdot 1^2 + 9 \cdot 10 \cdot 2^2 + 9 \cdot 5 \cdot 3^2 + 9 \cdot 3 \cdot 4^2 + 13 \cdot 10 \cdot 1^2 + 13 \cdot 5 \cdot 2^2 + 13 \cdot 3 \cdot 3^2 + \dots + 5 \cdot 3 \cdot 1^2} = 0.849$$

Most computations of  $\alpha$  can be performed with SPSS or SAS macros written by Andrew Hayes. Available at <http://www.afhayes.com>: Go to “SPSS and SAS Macros” then to “KALPHA.”

## References:

Klaus Krippendorff (2004). *Content Analysis, an Introduction to Its Methodology*, 2<sup>nd</sup> Edition. Thousand Oaks, CA: Sage Publications – especially Chapter 11, pages 211-256.

Klaus Krippendorff (2004). Reliability in Content Analysis: Some common Misconceptions and Recommendations. *Human Communication Research* 30,3: 411-433.  
[http://repository.upenn.edu/asc\\_papers/242](http://repository.upenn.edu/asc_papers/242) Last accessed 2011.1.25

Andrew F. Hayes, and Klaus Krippendorff (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures* 1: 77-89.  
<http://www.unc.edu/courses/2007fall/jomc/801/001/HayesAndKrippendorff.pdf> Last accessed 2011.1.25